

# Properties of the random seed input to Bell tests

Le Phuc Thinh,<sup>1</sup> Lana Sheridan,<sup>1</sup> and Valerio Scarani<sup>1,2</sup>

<sup>1</sup>*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, Singapore 117543*

<sup>2</sup>*Department of Physics, National University of Singapore, 3 Science Drive 2, Singapore 117542*

Device independent protocols rely on the violation of Bell inequalities to certify properties of the resources available. The violation of the inequalities are meaningless without a few well-known assumptions. One of these is measurement independence, the property that the source of the states measured in an inequality is uncorrelated from the measurements selected. Since this assumption cannot be confirmed, we consider the consequences of relaxing it and find that the definition chosen is critically important to the observed behavior. Considering a definition that is a bound on the min-entropy of the measurement settings, we find lower bounds on the min-entropy of the seed used to choose the inputs required to deduce any quantum or non-local behavior from a Bell inequality violation. These bounds are significantly more restrictive than the ones obtained by endowing the seed with the further structure of a Santha-Vazirani source. We also outline a procedure for finding tight bounds and study the set of probabilities that can result from relaxing measurement dependence.

PACS numbers: 03.67.-a, 03.65.Ta, 03.65.Ud

## I. INTRODUCTION

The violation of Bell inequalities can be used to certify important quantum information properties in a black-box scenario under minimal assumptions. This idea of “device-independent” certification started in the context of quantum key distribution, where the violation of Bell inequalities bounds the information leaked to the eavesdropper [1–3]; and it has been extended to various other tasks, notably state certification [1, 4, 5], measurement certification [6], and private randomness expansion [7–9]. Ultimately, this stems from the fact that the violation of Bell inequalities certifies the presence of a quantifiable amount of intrinsic randomness: indeed, *a contrario*, if the outcomes were predictable, one could have predicted them in advance and the measurement could consist of reading from a pre-existing list. This is exactly what the violation of Bell inequality certifies as impossible.

Two assumptions are left in device-independent certification. The first is *no-signaling*: the choice of the measurement setting of one party should not be known to the measurement boxes of the other parties before they produce their outcome. This can be guaranteed ultimately by ensuring space-like separation, although one may also trust a weaker demonstration of separation, as for instance in [7]. The second assumption is *measurement independence*: the information  $\lambda$  contained in the boxes in each run should be uncorrelated from the choice of the settings in that run. So far, no way of checking measurement independence is known in a black-box scenario: the best one can do is to buy the source of  $\lambda$  and the devices that choose the settings from different providers, who are believed not to be conspiring together. Alternatively, one can partly give up the black-box scenario, characterize the devices and be confident that the relevant degrees of freedom are uncorrelated.

It is clear that no-signaling and measurement independence cannot be arbitrarily relaxed: if any amount

of signaling is allowed, or if arbitrary correlation is admitted between source and settings, the violation of a Bell inequality can be obtained with purely classical resources  $\lambda$ , so there is no hope to conclude that  $\lambda$  contains intrinsic randomness. However, with *the aim of reducing the assumptions of device-independent certification to their bare minimum*, one can *partially* relax no-signaling and measurement independence, and ask how much information must be signaled and how much measurement dependence must be allowed for a Bell test to become irrelevant [10]. In this paper, we focus on the latter question, the study of *partial measurement dependence* (sometimes called reduced measurement independence or reduced “free will”), which has been the object of a few recent studies [11–14].

## II. MEASUREMENT DEPENDENCE AND ITS BASIC CONSEQUENCES

### A. Measurement independence

For the sake of this introduction, we consider a bipartite Bell scenario. Operationally, a Bell experiment consists of  $N$  apparently identical runs [32], in each of which box A receives input  $x$  and outputs a value  $a$ , box B receives input  $y$  and outputs a value  $b$ . One can then estimate the statistics  $p(ab|xy)$ . We denote by  $\lambda$  the information present in the boxes in a given run.

Measurement independence, the assumption that we want to relax, is captured by the condition

$$p(\lambda|xy) = p(\lambda) \quad \forall x, y. \quad (1)$$

Under this assumption, the observed statistics are modeled by

$$p_{\text{MI}}(ab|xy) = \int p(ab|xy\lambda)p(\lambda) d\lambda. \quad (2)$$

The specific goal of a Bell test is to assess whether there is intrinsic randomness in the boxes, that is, in the usual terminology, to guarantee that  $\lambda$  is not a *local variable*. Mathematically, local variables are defined by  $p(ab|xy\lambda) = p(a|x\lambda)p(b|y\lambda)$ . It is useful to stress that, as written, (2) contains an additional assumption, namely that  $\lambda$  itself is chosen independently in each run according to the distribution  $p(\lambda)$ . Under measurement independence, it can be proved that this is ultimately not a restriction for Bell tests, although one has to be careful in interpreting statistics from finite samples [15–17].

Measurement independence cannot be denied in a systematic way without undermining the scientific method itself (if a clinical trial is to make sense, whether each patient receives the drug or the placebo cannot depend on the any details of the patients’ conditions). However, it is certainly possible to question measurement independence in a given setup: the devices that determine the inputs  $x, y$  may be correlated to the process that determines  $\lambda$ . The origin of such correlation may be trivial, like the fluctuations in power of the city network to which all the devices are connected; it may be due to lack of attention of the experimentalists, who introduced unwanted connections; or it may be strongly conspiratorial, in an adversarial scenario in which the devices come from an untrusted provider. In all cases, (1) does not hold, nor does the proof that one can restrict the study to independently-chosen  $\lambda$ .

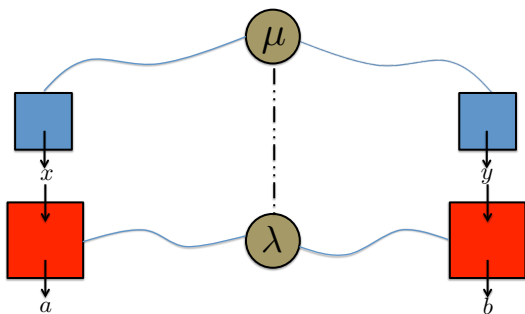


FIG. 1: (color online) There are many different processes by which the information  $\lambda$  in devices  $A$  and  $B$  might become correlated with the inputs to the devices  $x$  and  $y$ , as discussed in the text. In this illustration the processes are represented by some external pre-existing variable  $\mu$  that serves to introduce the correlation. The blue boxes represent the physical random number generators used to pick the inputs to the Bell test.

By relaxing condition (1), one allows correlations between the source  $\lambda$  and the choice of the settings  $x, y$ . Bayes theorem implies that

$$p(\lambda|xy) \neq p(\lambda) \iff p(xy|\lambda) \neq p(xy). \quad (3)$$

The first relation could be read as “the output of the source is restricted for a given choice of settings”, the second as “the choice of settings is restricted for a given output of the source”. Neither needs to refer to a real

causal relation: all is compatible with both  $\lambda$  and  $x, y$  being influenced by a common cause (Fig. 1). That being clarified, our discourse will be mostly phrased in the second way (the first way will be used in Section VI). We shall then look at measurement dependence as *reducing the probability of certain pairs of settings*. In the case where the dependence is sufficient to *exclude* enough pairs of settings, unwanted features of local variable models may be hidden. This is the same intuition behind the power of the detection loophole; in fact, measurement dependence is even stronger, because it may allow to exclude a single *pair* of settings, whereas the detection loophole is local and excludes all pairs of settings such that one given setting of (say) Bob is associated to unwanted features. This opens a wealth of possibilities that we review rapidly next.

## B. Effects of measurement dependence

The obvious effect of measurement independence is the possibility of *faking a violation of Bell inequalities*. A Bell inequality is built on a linear combination of  $p(ab|xy)$ , whose maximal value (called *algebraic limit*) cannot be reached by local variables. If, in each run, one can exclude some suitable pairs of settings in correlation with the content of the boxes  $\lambda$ , then it becomes possible to reach the algebraic limit while having only local variables in the boxes.

Let us illustrate this point with the most famous Bell inequality, that of Clauser, Horne, Shimony and Holt (CHSH). The inequality reads

$$|\langle a_0 b_0 \rangle + \langle a_0 b_1 \rangle + \langle a_1 b_0 \rangle - \langle a_1 b_1 \rangle| \leq 2 \quad (4)$$

with  $a_x, b_y \in \{-1, +1\}$ . In order to achieve the algebraic limit of 4, one should have  $a_0 = b_0$ ,  $a_1 = b_0$ ,  $a_0 = b_1$  and  $a_1 = -b_1$ . Local deterministic points exist that satisfy three out of these four conditions. If one wants to achieve the algebraic limit with local variable and measurement dependence, a sufficient strategy is the following: in each run,  $\lambda$  is chosen among the aforementioned local deterministic points, and the pair of settings corresponding to the unwanted condition is never chosen [10, 12].

The fact that a sufficient amount of measurement dependence can lead to the algebraic limit has an intriguing consequence for some inequalities. Indeed, in generic inequalities, the algebraic limit may lie even above what can be reached with no-signaling correlations. For instance, the tilted CHSH inequality

$$|\langle a_0 b_0 \rangle + \langle a_0 b_1 \rangle + \langle a_1 b_0 \rangle - \langle a_1 b_1 \rangle + \alpha \langle a_0 \rangle| \leq 2 + \alpha, \quad (5)$$

has an algebraic limit of  $4 + \alpha$ , but no-signaling correlations can reach only up to 4 if  $\alpha \leq 2$  [18]. If measurement dependence is allowed, to the point that one pair of settings can be excluded, then one can achieve the algebraic

limit with a convex mixture of

$$\begin{aligned} \lambda = (+1, -1, -1, +1) & \text{ together with } (x, y) \neq 00 \\ \lambda = (+1, +1, +1, -1) & \text{ together with } (x, y) \neq 01 \\ \lambda = (+1, -1, +1, +1) & \text{ together with } (x, y) \neq 10 \\ \lambda = (+1, +1, +1, +1) & \text{ together with } (x, y) \neq 11 \end{aligned} \quad (6)$$

where we denoted a local deterministic point as  $\lambda = (a_0, a_1, b_0, b_1)$ . If a Bell test is run with this underlying strategy, the observed correlations will lie outside the no-signalling polytope, i.e. are formally signalling. Obviously, this does not mean that measurement dependence makes it possible to use entanglement to actually send a message: in order for (say) Alice to send a message to Bob, she must be able to choose her setting at will, which is precisely what measurement dependence denies. At any rate, one must be careful when working with measurement dependence: the worst case are correlations that reach the algebraic limit, not the no-signaling one (to our knowledge, all the studies of measurement dependence so far dealt with inequalities for which the two limits happen to coincide [10–14]).

The take-away message of this paragraph is that one does not have to reach the extreme case of total measurement dependence (i.e.  $\lambda$  determining  $x, y$  uniquely): *already with some partial amount of measurement dependence, it becomes impossible to draw any conclusion from the violation of a Bell inequality*. One of our main result will consist in deriving general bounds for this amount (Section IV). In order to do that, we need first to define suitable figures of merit that quantify the amount of measurement dependence. This is the object of the following section.

### III. QUANTIFYING MEASUREMENT DEPENDENCE

From now onwards,  $p(xy|\lambda)$  will be called the *seed of the Bell test*. Partial measurement dependence means that there is some randomness in the seed, and we aim at quantifying this amount of randomness. The characterization of sources of randomness is a well studied subject in computer science, so we are going to adapt this known material to our problem [33]. Let us start by arguing that some attention is indeed required.

#### A. Critical choice of a figure of merit

We illustrate our point with an example. The *chained inequality* is a bipartite Bell inequality with  $m$  settings for each party and binary outcomes  $a, b$  for both measurements on  $A$  and  $B$ , which reads

$$\begin{aligned} I_m &= p(a = b | x = 1, y = m) + \sum_{\substack{x, y \text{ s.t.} \\ x \in \{y, y+1\}}} p(a \neq b | x, y) \\ &\leq 2m - 1. \end{aligned} \quad (7)$$

It has been used to put stringent bounds on quantum theory thanks to the property that, in the limit  $m \rightarrow \infty$ , its algebraic limit  $I_m = 2m$  can be reached with measurements on quantum states [19, 20].

Out of the  $m^2$  possible pairs of settings,  $2m$  are effectively used in the inequality. Furthermore, there exist local deterministic points that can satisfy  $2m - 1$  of these conditions. Therefore, in order to verify any conclusion based on the chained inequality, it is enough to have an amount of measurement dependence that allows the exclusion of only one pair of settings out of  $m^2$ . In the limit of large  $m$ , under whichever measure, such a seed is very close to a fully random source: for instance, its min-entropy per run (defined below) is  $\log(m^2 - 1)$ , which differs from the fully random value  $\log m^2$  by  $O(m^{-2})$ . This example shows that *a seed, which would presumably be considered as good as it gets in an abstract assessment, is already catastrophic for the Bell inequality under study*. Notice that this remark is not in contradiction with the results of [11], which can be seen as proving that the chained inequality is pretty robust to measurement dependence: indeed, in that work, the additional Santha-Vazirani assumption was made on the seed, which implies that all the pairs of settings are possible in each run. Our argument, based on excluding one setting in each run, does not apply.

It is now time to present the definitions we have just sketched in their suitable formal setting. We shall consistently use the word *seed* to stress that the source of randomness we are interested in is the randomness of the inputs given the knowledge of the physical process  $\lambda$  or vice versa, not the randomness possibly present in  $\lambda$  (which would be the intrinsic randomness of quantum origin in the ideal case).

#### B. Types of seed

Here we review rapidly some well-known types of sources of randomness for the purpose of this paper, referring to [21] for a comprehensive study.

Consider a random variable  $Z$  in an alphabet  $\mathcal{Z}$  of size  $d$ ; and let  $\mathbf{Z} = Z_1 \dots Z_N$  be an  $N$ -dit string. In our case,  $Z$  will represent the settings chosen for the Bell test, i.e.  $Z = (x, y)$  in a bipartite scenario. Randomness being synonymous with unpredictability, a source of randomness (here, the seed) will be characterized by specifying what one wants to predict and how predictable it is, given some prior information  $\Lambda$  (supposed to be classical throughout this paper). One would then say that the seed contains randomness if

$$P_{\text{guess}}(\mathbf{Z}|\Lambda) := \sum_{\lambda} P(\Lambda = \lambda) P_{\text{guess}}(\mathbf{Z}|\Lambda = \lambda) < 1, \quad (8)$$

where  $P_{\text{guess}}(\mathbf{Z}|\Lambda = \lambda) := \max_{\mathbf{z}} p(\mathbf{Z} = \mathbf{z} | \Lambda = \lambda)$ . The amount of randomness is quantified by the min-entropy

$$H_{\min}(\mathbf{Z}|\Lambda) := -\log P_{\text{guess}}(\mathbf{Z}|\Lambda). \quad (9)$$

Clearly,  $H_{\min}(\mathbf{Z}|\mathbf{\Lambda}) > 0$  implies the presence of some randomness. To someone who does not have access to  $\mathbf{\Lambda}$ , the source will appear to have min-entropy  $H_{\min}(\mathbf{Z}) = -\log P_{\text{guess}}(\mathbf{Z})$  which can only be higher by the data processing inequality [34].

The loosest characterization of the seed, i.e. the one that requires fewer assumptions, simply puts a bound on the min-entropy:

**Definition 1.** Min-entropy seed. *A random variable  $\mathbf{Z}$  is a  $k$ -min-entropy source of randomness with respect to another random variable  $\mathbf{\Lambda}$  if  $H_{\min}(\mathbf{Z}|\mathbf{\Lambda}) \geq k$ .*

As soon as  $k > 0$ , the knowledge of  $\mathbf{\Lambda}$  does not determine  $\mathbf{z}$  uniquely. One can add some structure to a min-entropy seed. For instance, a  $k$ -min-entropy seed is called *uniform* if  $H_{\min}(\mathbf{Z}|\mathbf{\Lambda} = \boldsymbol{\lambda}) := -\log P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda} = \boldsymbol{\lambda}) \geq k$  for all values of  $\boldsymbol{\lambda}$ . A *block min-entropy seed* is one for which not only the min-entropy of the whole string, but the min-entropy of blocks is also lower bounded. These notions will not be used in this paper.

As soon as  $k \leq \log(d^N - 1)$ , the definition of  $k$ -min-entropy seed is compatible with  $P_{\text{guess}}(\mathbf{Z} = \mathbf{z}|\mathbf{\Lambda} = \boldsymbol{\lambda}) = 0$  for one string  $\mathbf{z}$ . As hinted in paragraph III A, the possibility that some settings are not chosen is critical for seeds of Bell tests. Because of this, one may want to add to the properties of the seed the assumption that *all* the  $d^N$  strings have non-zero probability. This is equivalent to the following type of source:

**Definition 2.** Santha-Vazirani seeds. *A random variable  $\mathbf{Z}$  is a  $(p_{\min}, p_{\max})$  Santha-Vazirani source with respect to  $\mathbf{\Lambda}$  (where  $0 \leq p_{\min} \leq 1/d$  and  $1/d \leq p_{\max} \leq 1$ ) if*

$$p_{\min} \leq p(z_i|\boldsymbol{\lambda}, z_1, \dots, z_{i-1}) \leq p_{\max} \quad \forall i. \quad (10)$$

If  $Z_i$  is a bit,  $p_{\min} = 1 - p_{\max}$  is usually written  $\delta$  [22]. Some of the most important results in measurement dependence in Bell tests have been obtained for Santha-Vazirani sources [11, 13, 14]. These results show that there is a real advantage in considering Bell-based randomness, because it overcomes no-go theorems for classical information.

Finally, let us focus on distributions that are *independent and identically distributed (i.i.d.)* such that

$$p(\mathbf{Z} = \mathbf{z}|\mathbf{\Lambda} = \boldsymbol{\lambda}) = \prod_{j=1}^N p(Z_j = z_j|\lambda_j). \quad (11)$$

This can also be viewed as a block min-entropy source where each block consists of only one symbol,  $Z_i$ . In this case, the Santha-Vazirani definition implies:

$$p_{\min} \leq p(z|\lambda) \leq p_{\max}. \quad (12)$$

We will use a different notation such that  $p_{\max} = P_M$  and  $p_{\min} = P_m$  to make clear that we are in the i.i.d. scenario. Then the definition of uniform min-entropy

sources is equivalent to the figure of merit of measurement dependence used in [12], namely

$$P_M := \max_{z, \lambda} p(z|\lambda), \quad [\text{i.i.d.}] \quad (13)$$

since  $H_{\min}(Z|\Lambda = \lambda) \geq k$  for all  $\lambda$  is equivalent to  $P_M \leq 2^{-k}$ . In the following, we will use these two figure of merits interchangeably for i.i.d. models.

Instead of bounding the largest probability, the smallest probability also gives information on measurement dependence, as first proposed in [23]:

$$P_m := \min_{z, \lambda} p(z|\lambda). \quad [\text{i.i.d.}] \quad (14)$$

If only  $P_M$  is explicitly bounded, then a bound on  $P_m$  can be inferred, however, it might be trivial, since it can be negative:  $P_m \geq 1 - (d-1)P_M$ . Bounding only the min-entropy of the input seed to the Bell test, or equivalently bounding only  $P_M$ , which is the guessing probability, allows much different worst-case behavior in Bell tests than when the Santha-Vazirani definition is adopted, as we shall now explore.

#### IV. LOWER BOUND FOR MIN-ENTROPY SEEDS

We will be dealing with a  $K$ -partite Bell scenario where the  $i^{\text{th}}$  party has  $m_i > 1$  measurement settings ( $m_A, m_B$  for bipartite) and each setting has an arbitrary number of outcomes. The joint configuration of settings  $\mathbf{z} = z_1 \dots z_K$  with  $z_i \in \{1, \dots, m_i\}$  ( $\mathbf{z} = xy$  for bipartite) is a  $K$ -tuple in the set of all settings  $\mathcal{S}$  of size  $\prod_{i=1}^K m_i$ . In this Section, moreover, we consider a Bell test in which the observed statistics of the settings follow a uniform distribution, that is

$$H_{\min}(\mathbf{Z}_1, \dots, \mathbf{Z}_K) = N \log |\mathcal{S}|, \quad (15)$$

or equivalently

$$p_{\text{obs}}(z_1 \dots z_K) := \sum_{\lambda} p(z_1 \dots z_K|\lambda) p(\lambda) = \left( \prod_{i=1}^K m_i \right)^{-1}. \quad (16)$$

This is not an assumption like those on the nature of the seed:  $p_{\text{obs}}$  is observed in a realization; but it is a frequent working assumption for theoretical works, which was made in all previous works on measurement dependence. In Section V, we shall see that a non-uniform  $p_{\text{obs}}$  has interesting consequences in studies of measurement dependence.

We are presently able to discuss our main result: a lower bound on the min-entropy of the seed, below which no conclusion can be drawn from any Bell test, unless further structure is assumed.

### A. Reaching the no-signaling limit

The main insight is provided by the following Lemma, which we present in the bipartite scenario (the generalization to multipartite scenarios holds with identical proofs and more cumbersome notation, so we give it in Appendix A):

**Lemma 1.** *Let  $P(ab|xy)$  be an arbitrary no-signaling distribution with  $x \in \{1, \dots, m_A\}$  and  $y \in \{1, \dots, m_B\}$ . For any pair of settings  $(\bar{x}, \bar{y})$ , there exists a local distribution  $P_L(ab|xy)$  such that*

$$P_L(ab|xy) = P(ab|xy) \quad (17)$$

for  $(x, y) \in \mathcal{S}_{\bar{x}, \bar{y}} \equiv \{(\bar{x}, y'), (x', \bar{y}) : x' \in \{1, \dots, m_A\}, y' \in \{1, \dots, m_B\}\}.$

Moreover, this result is tight: if another pair of settings is added to the subset of pairs, there exists a no-signaling point for which those probabilities are nonlocal.

*Proof.* The proof can be done by constructing explicitly one such local distribution. Let us fix  $(\bar{x}, \bar{y}) = (1, 1)$  without loss of generality. From the no-signaling distribution  $P$ , we construct

$$\begin{aligned} \mathbf{P}(a_1, a_2, \dots, a_{m_A}; b_1, b_2, \dots, b_{m_B}) \\ = P(a_1)P(b_1|a_1) \prod_{j=2}^{m_A} P(a_j|b_1) \prod_{k=2}^{m_B} P(b_k|a_1) \end{aligned} \quad (18)$$

with obvious notations. This is a valid joint probability distribution over the outcomes of all the measurements. Now, on the one hand, the marginals  $\mathbf{P}(a_j; b_k) \equiv P_L(a, b|j, k)$  define a local distribution, as first proved by Fine [24]. On the other hand, it is easy to show that  $\mathbf{P}(a_1; b_k) = P(a, b|1, k)$ : one should sum first over all possible values of  $a_2, \dots, a_{m_A}$  to find  $\mathbf{P}(a_1; b_1, b_2, \dots, b_{m_B}) = P(a_1) \prod_{k=1}^{m_B} P(b_k|a_1)$ , after which the sum over the  $b$ 's is obvious. Similarly one proves that  $\mathbf{P}(a_j; b_1) = P(a, b|j, 1)$ . So indeed we have a local distribution that mimicks the initial no-signaling one on the desired subset of pairs of settings.

As for the tightness, suppose that we add a single pair of settings, say  $(2, 2)$ , to  $\mathcal{S}_{1,1}$ : there exist no-signaling points for which CHSH is violated by the settings  $(1, 1)$ ,  $(1, 2)$ ,  $(2, 1)$  and  $(2, 2)$ ; so those statistics can't be mimicked by a local distribution.  $\square$

Now we can state the main theorem:

**Theorem 1.** *Consider a min-entropy seed with an observed min-entropy  $H_{\min}(\mathbf{XY}) = N \log(m_A m_B)$  for an  $N$ -run bipartite Bell test with  $m_A$  inputs on Alice,  $m_B$  inputs on Bob and arbitrary alphabets for the outcomes. If*

$$H_{\min}(\mathbf{XY}|\mathbf{\Lambda}) \leq N \log(m_A + m_B - 1) \quad (19)$$

*no conclusion can be drawn from the Bell test, since the no-signaling limit of the inequality can be reached with*

*local distributions. The generalization of this result to  $K$ -partite Bell tests reads*

$$H_{\min}(\mathbf{Z}_1 \dots \mathbf{Z}_K | \mathbf{\Lambda}) \leq N \log \left( \sum_{k=1}^K m_k - K + 1 \right). \quad (20)$$

*Notice in particular that, without further assumptions, any source of randomness with  $H_{\min}(\mathbf{XY}|\mathbf{\Lambda}) \leq N \log 3$  is useless as a seed for any Bell tests.*

*Proof.* We will construct an explicit i.i.d. seed which allows the faking of a Bell violation up to the no-signalling bound with appropriate local resources. From Lemma 1 we know that there exist subsets  $\mathcal{S}_{\bar{x}, \bar{y}}$  of  $m_A + m_B - 1$  pairs of settings, for which no difference can be seen if a local distribution is substituted for a possibly nonlocal no-signaling point: in particular, this could be the no-signaling point that reaches the no-signaling limit for the inequality under study. If  $H_{\min}(\mathbf{XY}|\mathbf{\Lambda})$  is sufficiently low, the seed will allow only the pairs of settings that belong to one of the  $\mathcal{S}_{\bar{x}, \bar{y}}$  and distribute the corresponding local strategy  $\lambda_{\bar{x}, \bar{y}}$ . The seed

$$p(xy|\lambda_{\bar{x}, \bar{y}}) = \begin{cases} \frac{1}{m_A + m_B - 1}, & \text{if } x, y \in \mathcal{S}_{\bar{x}, \bar{y}} \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

has  $P_M = \frac{1}{m_A + m_B - 1}$  in each run, whence we have proved the bound (19) as long as we can find  $p(\lambda_{\bar{x}, \bar{y}})$  such that  $\sum_{\bar{x}, \bar{y}} p(xy|\lambda_{\bar{x}, \bar{y}}) p(\lambda_{\bar{x}, \bar{y}}) = p_{\text{obs}}(xy)$  for all  $x, y$ . In the case where  $p_{\text{obs}}$  is uniform, this can always be found by simply choosing uniformly the pair  $(\bar{x}, \bar{y})$ , i.e.  $p(\lambda_{\bar{x}, \bar{y}}) = \frac{1}{m_A m_B}$ . This concludes the proof for the bipartite case. The proof of the multipartite case is identical using the material of Appendix A. The final remark of Theorem 1 stems from the fact that each Bell test must involve at least two parties and each must have at least two settings.  $\square$

Because of the tightness of Lemma 1, the bounds (19) and (20) are the best *inequality-independent bounds* that one can obtain with i.i.d. seeds. Moreover, since there exist inequalities for which the quantum and the no-signaling limits coincide, the bound to reach the quantum limit cannot be better. If the inequality is given, however, much less measurement dependence may be sufficient to reach the no-signaling limit, and even less to reach the quantum limit if it is lower. We elaborate further on this point in the following paragraph.

### B. Inequality-dependent bounds

Let  $B$  define a Bell inequality, whose local, quantum and no-signaling limits are given by  $B_L \leq B_Q \leq B_{NS}$ , and  $\mathcal{S}^B$  be the set of settings that are used by the Bell inequality [35]. Again, for each  $\lambda$  there is a local strategy for assigning outputs such that in order to

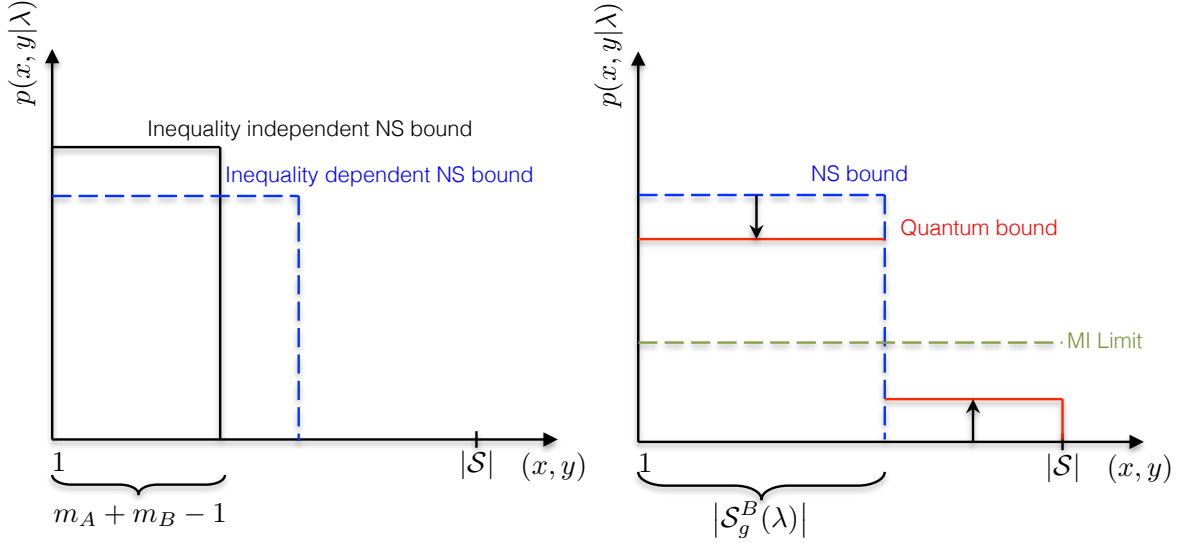


FIG. 2: (color online) Seeds that reach the critical min-entropy bound for uniform observed distribution of settings. For the inequality independent bound (20), the seed is uniform on  $m_A + m_B - 1$  settings and is zero elsewhere. For a given inequality, the no-signaling limit may be reached with a seed that is uniform on a larger number of settings  $|\mathcal{S}_g^B|$ , and still zero on the others; in order to reach only the quantum limit, one can allow the settings in  $\mathcal{S}_h^B$  to be used sometimes. Of course, for each  $\lambda$ , the settings that are chosen may vary.

achieve the no-signaling limit, some settings will be incompatible with this strategy and must be hidden by measurement dependence. Let this set of inputs be  $\mathcal{S}_h^B(\lambda)$ . Then, an arbitrary no-signaling point is required to be compatible with a local point only on the subset  $\mathcal{S}_g^B(\lambda) := \mathcal{S}^B \setminus \mathcal{S}_h^B(\lambda)$ . Suppose an inspection of the inequality  $B$  shows that *at most*  $|\mathcal{S}_h^B|$  of these  $|\mathcal{S}^B|$  settings must be hidden for any choice of  $\lambda$ . Once the probabilities of the settings  $\mathcal{S}_h^B(\lambda)$  are set to zero, the min-entropy is maximized by the uniform distribution over the remaining  $|\mathcal{S}_g^B|$  settings (FIG. 2). However, one must be very careful to show the existence of  $p(\lambda)$  which satisfies (16). Whenever such a distribution exists, if  $P_M \geq 1/|\mathcal{S}_g^B|$ , the non-local game can be won with probability one with local strategies. As implied by the results of the previous section, if the observed input distribution  $p_{obs}(\mathbf{z})$  is uniform then a strategy in the form of equation (21) or its generalization in Appendix A with a uniform probability over  $\lambda$  will always satisfy (16). However, it is possible to do better in some cases where  $|\mathcal{S}^B| < |\mathcal{S}|$ . In such cases  $|\mathcal{S}_h^B|$  (the most settings that must be hidden for any  $\lambda$ ) can be small. Here, for a uniform  $p_{obs}$  equation (16) will also be satisfied provided possibly more settings than required are hidden for each  $\lambda$  such that  $|\mathcal{S}_h^B(\lambda)| = |\mathcal{S}_h^B|$  for all  $\lambda$  and if  $\mathcal{L}(\mathbf{z})$  is the set of  $\lambda$ s for which  $\mathbf{z} \in \mathcal{S}_h^B(\lambda)$ ,  $|\mathcal{L}(\mathbf{z})| = |\mathcal{S}_h^B|$  must be constant for all  $\mathbf{z}$ . This is a symmetry condition that can be met by many Bell inequalities. As before, the existence of this example proves that a min-entropy seed with

$$k \leq N \log(|\mathcal{S}_g^B|) \quad (22)$$

can reach the no-signaling limit of  $B$  with local strategies for uniform input distributions. In the following section

we will show how to obtain bounds for arbitrary  $p_{obs}$  and that approach will also give tight bounds and optimal strategies when the inequality is one in which the size of the “hidden sets” varies with  $\lambda$ .

Further, if  $B_Q < B_{NS}$ , in order to simulate physics one may be content with *reaching the quantum limit*. A possible i.i.d. seed (not proved to be optimal) is the following (see Fig. 2). With probability  $1 - q$ , the settings are chosen uniformly among all  $\mathcal{M}$  possible  $K$ -tuples: this is measurement independence, so  $B \leq B_L$  on these cases, and the physical process  $\lambda$  can be chosen as one of those that saturate  $B = B_L$ . In the other instances, the settings are chosen uniformly in  $\mathcal{S}_g^B$  and the physical process  $\lambda$  is chosen in each case in order to achieve  $B = B_{NS}$ . In other words, this seed is a convex combination of the measurement independent uniform seed and the seed described in the previous paragraph. Note that this new seed will automatically satisfy the constraint (16). For such a seed, therefore,  $P_M$  is the probability of each setting in  $\mathcal{S}_g^B$ , which reads  $P_M = \frac{1}{|\mathcal{S}|} [1 + q(|\mathcal{S}|/|\mathcal{S}_g^B| - 1)]$ . With this measurement-dependent strategy, one can reach  $B = qB_{NS} + (1 - q)B_L$ , so  $B \geq B_Q$  for  $q \geq (B_Q - B_L)/(B_{NS} - B_L)$ . In summary, the quantum limit can be achieved with an i.i.d. seed with

$$P_M^{B,Q} \geq \frac{1}{|\mathcal{S}|} \left[ 1 + \frac{B_Q - B_L}{B_{NS} - B_L} (|\mathcal{S}|/|\mathcal{S}_g^B| - 1) \right], \quad (23)$$

that is, a min-entropy seed with  $k \leq -N \log P_M^{B,Q}$  can reach the quantum limit of  $B$  with local strategies, for a uniform input distribution.

Let us illustrate the methodology with the analysis of some inequalities:

- *CHSH*: here, it is always necessary and sufficient to hide one pair of settings. Therefore  $|\mathcal{S}_g^B| = 3$  and the inequality-dependent bound (22) is the same as the inequality-independent one (19) to reach the no-signaling limit, as already proved in [12]. Recall that this does not prove the bounds to be tight, because they are based on explicit i.i.d. seeds: non i.i.d. seeds may lead to tighter bounds, though we do not know any example. As for reaching the quantum limit, we have  $P_M^{B,Q} = \frac{1}{4}[1 + (\sqrt{2}-1)/3] \approx 0.2845$ .
- *Chained inequality*: here again, as we have seen in paragraph III A, it is always necessary and sufficient to hide only one pair of settings out of  $\mathcal{M} = m^2$ , so  $|\mathcal{S}_g^B| = m^2 - 1$  and  $|S_h^B(\lambda)| = 1$  for all  $\lambda$ . As a consequence, in terms of min-entropy, the inequality-dependent bound (22) is  $N \log(m^2 - 1)$ , which is approximately twice the value  $N \log(2m - 1)$  obtained from (19). For large  $m$ , the quantum and no-signaling limits basically coincide.
- *CGLMP inequalities*: like the CHSH inequality, the CGLMP inequalities are two party inequalities where each party has two inputs. However, this family of inequalities has  $d$  possible outputs for each party. In the quantum case, the CGLMP inequalities can provide more robustness against measurement dependence than the CHSH inequality, in the sense that the min-entropy of the inputs given the source must be lower if the quantum bound is to be achieved. The reason is that it has been shown that as  $d \rightarrow \infty$ , the quantum limit increases and approaches the no-signalling limit [25, 26]. As can be seen, inspecting equation (23), the value of  $(B_Q - B_L)/(B_{NS} - B_L)$  will increase with  $d$ , and the value of  $P_M$  necessary to reach the quantum limit with local resources increases, until it reaches the no-signalling value  $P_M^{B,NS}$  in the limit.

## V. THE POSITIVE EFFECT OF BIASING THE CHOICES OF THE SETTINGS

Theorem 1 shows that assuming a full min-entropy source on the measurement settings, for any meaningful conclusion to be drawn from a Bell test, it must be that  $H_{\min}(\mathbf{XY}|\mathbf{\Lambda}) > N \log(m_A + m_B - 1)$ . However, recalling that the role of the observed data is actually a *constraint* imposed on the underlying model (similar to equation (16)), we can hope to use it to our advantage. This motivates the question: for a given value of  $H_{\min}(\mathbf{XY}|\mathbf{\Lambda}) = Nk > N \log(m_A + m_B - 1)$  that is being assumed, what is the optimal distribution on the inputs such that the maximum possible Bell value obtainable with this degree of measurement dependence and only local resources is as low as possible. Because the situation for non i.i.d. models is intractable, we are restrict-

ing ourselves to the i.i.d. model for the remaining of this chapter. Here instead of the min-entropy, the guessing probability  $P_M$  is used exclusively as the figure of merit of measurement dependence. First, we consider the CHSH inequality as an explicit example.

### A. The CHSH Inequality

Intuitively, we expect that the optimal solution is to set for each input round  $H_{\min}(XY) = H_{\min}(XY|\mathbf{\Lambda}) = k$  and  $p_{obs}(xy) = 2^{-k}$  for three pairs  $(x, y)$  and  $p_{obs}(x'y') = 1 - 3 \times 2^{-k}$  for the final pair because in this case  $\mathbf{\Lambda}$  cannot contain any further information on  $XY$  than is available simply from observing the distribution  $p_{obs}(xy)$ . We will highlight an example of this type of distribution later in this section. This is not a uniform distribution, so we can already see that non-uniform input distributions can be beneficial. In this section, we will consider fixed input distributions  $p_{obs}(xy)$  and find the maximum value that the CHSH inequality can take given a bound on  $P_M$ . Note that the method in this section extends to any multipartite Bell inequality.

We want to find the violation  $B_{CHSH}^{\max}$ , under local resources and measurement dependence, as a function of  $P_M$  and  $p_{obs}(xy)$ . To this end, observe that the local distributions form a convex polytope and so is the set of seeds with a fixed value of  $P_M$  (the seed polytope). Using the decomposition into extremal points of a convex polytope, we have

$$p(ab|xy\lambda) = \sum_i \alpha_i(\lambda) e_i(ab|xy), \quad (24)$$

$$p(xy|\lambda) = \sum_j \beta_j(\lambda) f_j(xy), \quad (25)$$

where  $e_i(ab|xy)$  are the extremal points of the local polytope and  $f_j(xy)$  are the extremal points of the seed polytope. Now after multiplying by both sides by  $p_{obs}(xy)$ , the i.i.d. model with measurement dependence becomes

$$\begin{aligned} p(abxy) &= \int_{\Lambda} d\lambda \sum_{ij} \alpha_i(\lambda) \beta_j(\lambda) e_i(ab|xy) f_j(xy) p(\lambda) \\ &= \sum_{ij} \gamma_{ij} g_{ij}(abxy), \end{aligned} \quad (26)$$

where

$$\gamma_{ij} = \int_{\Lambda} d\lambda \alpha_i(\lambda) \beta_j(\lambda) p(\lambda), \quad (27)$$

$$g_{ij}(abxy) = e_i(ab|xy) f_j(xy). \quad (28)$$

In this notation, the problem becomes a linear program, i.e. finding

$$B_{CHSH}^{\max}(P_M, p_{obs}(xy)) = \max_{p(abxy)} \sum_{abxy} (-1)^{a+b+xy} \frac{p(abxy)}{p_{obs}(xy)} \quad (29)$$



subjected to the constraints

$$p(abxy) = \sum_{ij} \gamma_{ij} g_{ij}(abxy), \quad \sum_{ab} p(abxy) = p_{obs}(xy) \quad (30)$$

for known values of  $p_{obs}(xy)$  and  $P_M$ . The result is presented in FIG. 3.

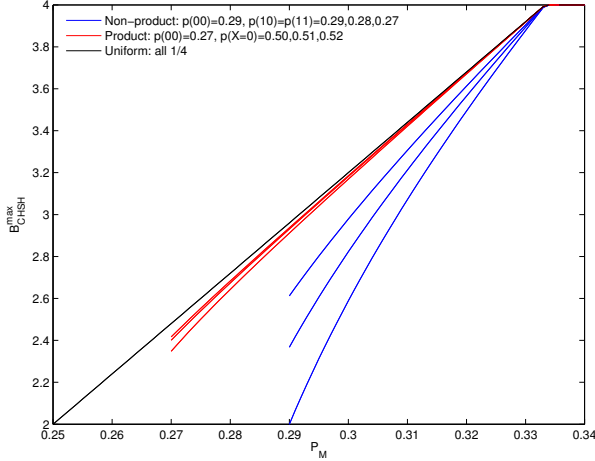


FIG. 3: (color online) Plot of maximum CHSH value against measurement dependence  $P_M$  for different  $p_{obs}$ . Notice that  $P_M$  start from  $\max\{p_{00}, p_{01}, p_{10}, p_{11}\}$  because of the data processing inequality: no underlying model of smaller  $P_M$  can reproduce the observed input statistics. In a Bell test with an assumed dependence bound  $P_M$ , if the value of the inequality is above the line that corresponds to the observed input distribution  $p_{obs}$  then there is intrinsic randomness in the outcomes contributed by  $\lambda$ . Therefore, for some observed violations, biased settings statistics can allow the certification of intrinsic randomness while uniform statistics cannot.

Using the numerical results, it is easy to see that the optimal strategy for maximizing the Bell value  $B_{CHSH}$  whether or not the observed distribution is uniform is to choose

$$p(xy|\lambda_{\bar{x},\bar{y}}) = \begin{cases} P_M & \text{if } x, y \in \mathcal{S}_{\bar{x},\bar{y}} \\ 1 - 3P_M & \text{otherwise} \end{cases}, \quad (31)$$

for each  $\lambda_{\bar{x},\bar{y}}$ , where  $\mathcal{S}_{\bar{x},\bar{y}}$  is defined in equation (18). Choosing this strategy, it is straightforward to find an analytic expression for  $B_{CHSH}^{\max}$ :

$$B_{CHSH}^{\max}(P_M) = 4 - \frac{1}{2} \left[ (1 - 3P_M)q + \frac{1 - 3P_M}{4P_M - 1}(q - 16) \right]$$

where for convenience we define  $q := \sum_{x,y} \frac{1}{p_{obs}(xy)}$ . This expression is only valid for  $\max_{x,y} p_{obs}(xy) \leq P_M < \frac{1}{3}$  ( $H_{\min}^{obs}(xy) \geq H_{\min}(xy|\lambda) > \log 3$ ). Notice that when the distribution is uniform  $q = 16$  and the second term vanishes, leaving a linear expression in  $P_M$ .

It is interesting to observe that for the purpose of violating Bell inequalities (that is, demonstrating non-locality by exceeding  $B^{\max}$ ) under measurement dependence, suppose the inputs have privacy quantified by  $P_M$ ,

then it is advantageous for us to *purposely* select an input distribution that is not uniform. This can be seen easily from for example the red curves for  $P_M = 0.27$ : selecting uniform input distribution allows a violation up to about 2.5 while selecting non-uniform input distribution only allows a lower maximum violation! Note that for non-uniform distributions on the inputs the upper bound on the Bell value is only as low as 2 (the local bound assuming measurement independence) for non-product distributions on the inputs. (See the blue curves.) All non-uniform product input distributions can have Bell values larger than 2, if measurement independence is relaxed. Notice also, that the lowest blue curve, the one that takes the value 2 at  $P_M = 0.29$  is the one corresponding to the distribution  $[p(00), p(01), p(10), p(11)] = [0.29, 0.13, 0.29, 0.29]$ . This is precisely the form of the distribution on the inputs we have anticipated at the start of this section.

## B. Generalizations

We have seen that for the case of the CHSH inequality, the strategy outlined in section IV in equation (21) is the optimal strategy even in the case that the distribution  $p_{obs}(xy)$  is not uniform. In general however this is not the case. It is possible to find some inequalities that together with some distributions  $p_{obs}(\mathbf{z})$  do not admit a strategy of the form

$$p(xy|\lambda_{\bar{\mathbf{z}}}) = \begin{cases} P_M, & \text{if } \mathbf{z} \in \mathcal{S}_{\bar{\mathbf{z}}} \\ Q(\lambda_{\bar{\mathbf{z}}}), & \text{otherwise} \end{cases} \quad (32)$$

where  $Q(\lambda)$  is determined by the normalization condition to be  $Q(\lambda_{\bar{\mathbf{z}}}) = \frac{1 - |S_g^B(\lambda)|}{|S_h^B(\lambda)|}$ .

Let us limit our focus to inequalities with symmetries such that  $|S_g^B(\lambda)| = |S_g^B|$  and  $|S_h^B(\lambda)| = |S_h^B|$  for all  $\lambda$ . In that case, equation (16) can be written as a matrix equation, with  $p_{obs}(\mathbf{z})$  and  $p(\lambda)$  written as vectors and  $p(\mathbf{z}|\lambda_{\bar{x},\bar{y}})$  is a matrix whose entries are defined by equation (32). If the  $p(\mathbf{z}|\lambda)$  matrix is full-rank, then there is a unique solution for  $p(\lambda)$  that is a valid probability distribution. This will always be the case if  $|S_g|$  and  $|S_h|$  have no common factors.

Examples of cases where the sizes of the sets  $S_g$  and  $S_h$  have no common factors are any bipartite Bell inequality with terms for all input pairs present and where both parties have the same number of inputs. For these cases, the min-entropy bound of section IV also applies for any non-uniform observed distribution on the inputs.

If, for a given inequality,  $|S_g|$  and  $|S_h|$  have at least one common prime factor, there may be some choices of distribution  $p_{obs}$  for which the strategy (32) will not be able to reproduce  $p_{obs}$  with any valid distribution  $p(\lambda)$ . In that case, the optimal strategy may have to be found numerically. For the i.i.d. case, one do this by solving a linear program that is a generalization of the one presented in the previous section.



## VI. BOUNDS ON THE ACHIEVABLE DISTRIBUTIONS

While the set of distributions obtainable from a measurement independence local model is the local polytope, that obtainable from a measurement dependence model is in principle a larger set. The example in section II B makes it clear that this set can even include signalling distributions. Now we wish to study more carefully this new set of distributions. As mentioned in II A, figures of merit of measurement dependence can be defined as restrictions on  $p(\lambda|xy)$  instead of  $p(xy|\lambda)$ . It turns out that in characterizing the set of achievable distributions, it is easier to work with figures of merits based on  $p(\lambda|xy)$ . In general, specifying  $p(xy|\lambda)$  does not specify  $p(\lambda|xy)$  unless  $p(xy)$  and  $p(\lambda)$  are uniform, in which case the two are proportional. Among the figures of merit one can define is the one by Hall [10]

**Definition 3.** Hall’s Measurement Dependence. *The quantity  $M$  quantifies the change to the probability distribution over a random variable  $\Lambda$  given  $X = x, Y = y$  versus the distribution given  $X = x', Y = y'$ :*

$$M := \max_{x, x', y, y'} \int |p(\lambda|x, y) - p(\lambda|x', y')| d\lambda. \quad (33)$$

$M$  can range from 0, the measurement independence case, to 2, corresponding to the probability of a given value of  $\lambda$  being zero for at least one of the input pairs  $(x, y)$  and  $(x', y')$ .

An alternative figure of merit that we will use in this section is the following

**Definition 4.** *The quantity  $M'$  bounds the distance between the probability distribution over the random variable  $\Lambda$  given access to  $x, y$  and the distribution over  $\Lambda$  without information on  $x, y$ :*

$$M' := \max_{x, y} 2D(p(\Lambda|X = x, Y = y), p(\Lambda)), \quad (34)$$

with  $p(\Lambda) = \sum_{x, y} p(\Lambda|X = x, Y = y)p_{\text{obs}}(X = x, Y = y)$  and  $D(\cdot, \cdot)$  is the total variational distance.

Essentially this definition bounds the distance of any one element in the distribution away from a strategy independent of Alice and Bob’s inputs  $x$  and  $y$ , which is similar in flavour to a Santha-Vazirani bound.

Now we wish to compare the point obtainable from a measurement dependence model with local resources,

$$p_{ab|xy}^\lambda = \sum_{\lambda} p(a|x\lambda)p(b|y\lambda)p(\lambda|xy), \quad (35)$$

to its corresponding measurement independent point,

$$p_{ab|xy} = \sum_{\lambda} p(a|x\lambda)p(b|y\lambda)p(\lambda). \quad (36)$$

Their distinguishability is characterized by their total variational distance

$$\begin{aligned} D(p_{AB|xy}^\lambda, p_{AB|xy}) &= \frac{1}{2} \sum_{a, b} |p_{ab|xy}^\lambda - p_{ab|xy}| \\ &= \frac{1}{2} \sum_{a, b} \left| \sum_{\lambda} p(a|x\lambda)p(b|y\lambda)(p(\lambda|xy) - p(\lambda)) \right| \\ &\leq \frac{1}{2} \sum_{\lambda} \left( \sum_{a, b} p(a|x\lambda)p(b|y\lambda) \right) |p(\lambda|xy) - p(\lambda)| \\ &= D(p(\lambda|x, y), p(\lambda)) \\ &\leq M'/2, \end{aligned} \quad (37)$$

where we have applied the triangle inequality. In other words,

$$|p_{ab|xy}^\lambda - p_{ab|xy}| \leq M' \quad \forall a, b, x, y, \quad (38)$$

which gives a bound on the distributions that can be created with a measurement dependence model of dependence bounded by  $M'$ .

There is another model that relaxes assumptions about correlations between two parties: quantum “cross-talk” between non-fully-isolated devices [27]. In this case, the assumption that the measurement operators are in a tensor product,  $M_A \otimes M_B$ , is relaxed. Of course in this case, it is possible to use the measurement itself to introduce quantum correlations between the two parties. In [27] it is assumed that the source of the measured state distributed to Alice and Bob is not subject to the cross-talk. The model was proposed to describe excess correlations that could result from a pair of trapped ions measured while situated next to each other in a refrigerator after being entangled.

As introduced in [27], let us consider the cross-talk model bounded by  $\chi$

$$-\chi \mathbb{I} \preceq \inf_{\Pi_A, \Pi_B} (\Pi_{AB} - \Pi_A \otimes \Pi_B) \preceq \chi \mathbb{I}, \quad (39)$$

where  $\Pi_A, \Pi_B$ , and  $\Pi_{AB}$  are POVM elements. As noted in [27], the bound (39) implies

$$|p'_{ab|xy} - p_{ab|xy}| \leq \chi, \quad (40)$$

where the prime indicates the distribution resulting from non-zero cross-talk. Comparing this with equation (38), it is clear that a cross-talk model bounded by  $\chi$  implies the same bounds on the deviation of the distributions using the extra correlations (cross-talk) as the measurement dependent model when bounded by  $M' = \chi$ .

## VII. CONCLUSIONS

Bell tests are an essential tool in device-independent approaches. They rely on a set of reasonable assumptions, but some of the assumptions are untestable. In

particular, the correlations between source and settings are *strictly unobservable* and therefore the amount of reduction of measurement independence is ultimately an assumption, either on the power of an adversary, on a physical model for the experiment. This study has demonstrated that when relaxing this assumption, the definition used, be it min-entropy or a Santha-Vazirani condition, is critical with respect to what kind of guarantees can be obtained from a Bell test. There are results [13, 28] showing that with a Santha-Vazirani source assumption arbitrarily weak randomness can be amplified using a protocol that checks for the violation of a Bell inequality. This cannot be accomplished using a min-entropy condition, as we have demonstrated in section IV: for sufficiently low min-entropy any inequality can be violated up to its no-signalling bound, using only the classical measurement dependent correlations and in a way that a third party could predict all of the outcomes of the measurements. Even for the protocol in [11] that amplifies bounded randomness (in the Santha-Vazirani definition) using violations of the chained Bell inequality, in order to get perfectly free bits out, the number of outputs for this inequality must go to infinity. As we point out in section III A, in this limit the chained Bell inequality is not robust to any relaxation of input randomness if the min-entropy definition is used instead.

The bounds on the min-entropy presented in section IV give immediate bounds for any inequality on the amount of input randomness required to draw conclusions about whether the violation of a Bell inequality can give any certification of quantum or non-local behavior. The method demonstrated for the CHSH inequality in sec-

tion V demonstrates how to get tight upper bounds for the value a given Bell inequality can take assuming a min-entropy bound for any distribution over the measurement settings. It also shows that there may be advantages to deliberately choosing non-uniform distributions over measurement settings in device independent protocols, depending on what assumptions are being made. Relaxing the assumption of measurement-dependence increases the set of probability distributions  $\{p(ab|xy)\}$  that can result from a Bell test assuming a local, realistic hidden variable model and section VI gives an expression bounding this increase.

Being as the assumption of measurement independence cannot be confirmed, it is important to understand the consequences for device independent protocols when it is relaxed. It is especially interesting that the min-entropy condition, a condition widely adopted in classical security studies [21, 29–31], has such a different behavior from the Santha-Vazirani condition for these device-testing purposes. We hope that the bounds and characterizations provided here will be useful for constructing protocols that are more robust to extraneous correlations.

#### Acknowledgments

This work is funded by the Centre for Quantum Technologies, which is funded by the Singapore Ministry of Education and the Singapore National Research Foundation. We would like to thank Jean-Daniel Bancal, Jeysthur Ang, Roger Colbeck and Yaoyun Shi for helpful discussions.

- 
- [1] D. Mayers and A. Yao. *Proceedings of the 39th IEEE Conference on Foundations of Computer Science*, page 503, 1998.
  - [2] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani. *Phys. Rev. Lett.*, 98:230501, 2007.
  - [3] U. Vazirani and T. Vidick. 2012. arXiv:1210.1810.
  - [4] C.E. Bardyn, T.C.H. Liew, S. Massar, M. McKague, and V. Scarani. *Phys. Rev. A*, 80:062327, 2009.
  - [5] M McKague, T H Yang, and V Scarani. Robust self-testing of the singlet. *Journal of Physics A: Mathematical and Theoretical*, 45(45):455304, 2012.
  - [6] R. Rabelo, M. Ho, D. Cavalcanti, N. Brunner, and V. Scarani. *Phys. Rev. Lett.*, 107:050502, 2011.
  - [7] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe. *Nature*, 464:1021, 2010.
  - [8] R. Colbeck and A. Kent. *J. Phys. A: Math. Theor.*, 44:095305, 2011.
  - [9] U. Vazirani and T. Vidick. 2012. arXiv:1111.6054.
  - [10] M. J. W. Hall. *Phys. Rev. A*, 84:022102, 2011.
  - [11] R. Colbeck and R. Renner. *Nature Physics*, 8:450, 2012.
  - [12] Dax Enshan Koh, Michael J. W. Hall, Setiawan, James E. Popen, Chiara Marletto, Alastair Kay, Valerio Scarani, and Artur Ekert. Effects of reduced measurement independence on bell-based randomness expansion. *Phys. Rev. Lett.*, 109:160404, Oct 2012.
  - [13] R. Gallego, L. Masanes, G. De La Torre, C. Dhara, L. Aolita, and A. Acín. 2012. arXiv:1210.6514.
  - [14] P. Mironowicz and M. Pawłowski. 2013. arXiv:1301.7722.
  - [15] J. Barrett, D. Collins, L. Hardy, A. Kent, and S. Popescu. *Phys. Rev. A*, 66:042111, 2002.
  - [16] R.D. Gill. *Mathematical Statistics and Applications: Festschrift for Constance van Eeden*. Eds: M. Moore, S. Froda and C. Léger. *IMS Lecture Notes – Monograph Series*, 42:133, 2003.
  - [17] Y. Zhang, S. Glancy, and E. Knill. *Phys. Rev. A*, 84:062118, 2011.
  - [18] A. Acín, S. Massar, and S. Pironio. *Phys. Rev. Lett.*, 108:100402, 2012.
  - [19] J. Barrett, A. Kent, and S. Pironio. *Phys. Rev. Lett.*, 97:170409, 2006.
  - [20] R. Colbeck and R. Renner. *Phys. Rev. Lett.*, 101:050403, 2008.
  - [21] S. Vadhan. *Pseudorandomness*. Foundations and Trends in Theoretical Computer Science. 2012.
  - [22] M. Sántha and U. V. Vazirani. *Journal of Computer and System Sciences*, 33:75, 1984.

- [23] M. Pawłowski, K. Horodecki, P. Horodecki, and R. Horodecki. 2009. arXiv:0902.2162.
- [24] Arthur Fine. Hidden variables, joint probability, and the bell inequalities. *Phys. Rev. Lett.*, 48:291–295, Feb 1982.
- [25] S. Zohren and R. Gill. *Phys. Rev. Lett.*, 100:120406, 2008.
- [26] S. Zohren, P. Reska, R. Gill, and W. Westra. *Europhysics Lett.*, 90:10002, 2010.
- [27] J. Silman, S. Pironio, and S. Massar. 2012. arXiv:1211.5921.
- [28] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, M. Pawłowski, and R. Ramanathan. 2013. arXiv:1303.5591.
- [29] R. Impagliazzo, L. Levin, and M. Luby. In *STOC '89 Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 12–24, 1989.
- [30] Y. Dodis and J. Spencer. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 376 – 385, 2002.
- [31] Y. Dodis, S. J. Ong, M. Prabhakaran, and A. Sahai. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 196 – 205, 2004.
- [32] We focus on the operational description of current experiments and do not consider the more general, but as yet abstract, case of *parallel repetition*, in which all the inputs are given at the same time.
- [33] Notice that different criteria, directly inspired by the problem of measurement dependence in a Bell test, have been put forward in [10].
- [34] Though obvious, it may be worth stressing that  $\sum_{\lambda} P(\mathbf{\Lambda} = \lambda) P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda} = \lambda)$  is *not* the same as  $P_{\text{guess}}(\mathbf{Z})$ , since  $P_{\text{guess}}$  is not a given probability distribution but a notation for a procedure that picks up the maximum of a probability distribution. As an extreme example, if  $\mathbf{Z}$  looks uniform but the knowledge of  $\mathbf{\Lambda}$  determines  $\mathbf{z}$  uniquely, one has  $P_{\text{guess}}(\mathbf{Z}) = \frac{1}{d^N}$  and

- $\sum_{\lambda} P(\mathbf{\Lambda} = \lambda) P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda} = \lambda) = 1$ .
- [35] The chained Bell inequality is an example of an inequality whose value depends only some of the possible inputs. Only terms such that  $x = y$  or  $x = y + 1$  and the term for  $x = 1, y = m$  appear.

## Appendix A: Generalization of Lemma 1 the multipartite case

The bound (20) for the min-entropy in a multipartite scenario is based on the generalization of Lemma 1 that we provide here:

**Lemma 2.** *Let  $P(\mathbf{o}|\mathbf{z})$  be an arbitrary  $K$ -partite no-signaling distribution with  $\mathbf{z} = (z_1, \dots, z_K)$  where  $z_i \in \{1, \dots, m_i\}$  and  $\mathbf{o}$  is a  $K$ -tuple of outcomes. For any  $K$ -tuple of settings  $\bar{\mathbf{z}} = (\bar{z}_1, \dots, \bar{z}_K)$ , there exists a local distribution  $P_L(\mathbf{o}|\mathbf{z})$  such that*

$$P_L(\mathbf{o}|\mathbf{z}) = P(\mathbf{o}|\mathbf{z}) \quad (\text{A1})$$

for  $\mathbf{z} \in \mathcal{S}_{\bar{\mathbf{z}}} \equiv \{(\bar{z}_1, z'_2, \dots, z'_K), \dots, (z'_1, \dots, z'_{K-1}, \bar{z}_K) : z'_i \in \{1, \dots, m_i\}\}$ .

*Moreover, this result is tight: if another  $K$ -tuple of settings is added to the subset  $\mathcal{S}_{\bar{\mathbf{z}}}$ , there exist a no-signaling point for which those probabilities are nonlocal.*

*Proof.* Again, let us fix  $\bar{\mathbf{z}} = (1, \dots, 1)$  without loss of generality. Let  $o_{z_i}^i$  be the  $i$ th party's outcome given the  $z_i$ th measurement setting. From the no-signaling distribution  $P$ , we construct a valid probability distribution

$$\mathbf{P}(o_1^1 \dots o_{m_1}^1; \dots; o_1^K \dots o_{m_K}^K) = P(o_1^1) \left[ \prod_{i=2}^K P(o_1^i | o_1^1 \dots o_1^{i-1}) \right] \left[ \prod_{i=1}^K \prod_{j=2}^{m_i} P(o_j^i | o_1^1 \dots o_1^{i-1} o_1^{i+1} \dots o_1^K) \right] \quad (\text{A2})$$

whose marginals  $\mathbf{P}(o_{z_1}^1; \dots; o_{z_K}^K) \equiv P_L(o_1^1 \dots o_K^K | z_1 \dots z_K)$  define a local distribution by Fine's result [24]. To verify that we have a local distribution that mimics the initial no-signaling one on the desired subset of pairs of settings, consider this example: for the input string

$(1, z_2, \dots, z_K)$  with the distribution  $\mathbf{P}(o_1^1; o_{z_2}^2; \dots; o_{z_K}^K) = P_L(o_1^1 o^2 \dots o^K | 1, z_2, \dots, z_K)$  we sum first over all possible values of each outcome variable  $o_2^1, \dots, o_{m_1}^1$  to find

$$\mathbf{P}(o_1^1; o_1^2 \dots o_{m_2}^2; \dots; o_1^K \dots o_{m_K}^K) = P(o_1^1) \prod_{i=2}^K \left[ P(o_1^i | o_1^1 \dots o_1^{i-1}) \prod_{j=2}^{m_i} P(o_j^i | o_1^1 \dots o_1^{i-1} o_1^{i+1} \dots o_1^K) \right] \quad (\text{A3})$$

after which continue to sum over all the  $o_k^2, \dots, o_k^K$  except

$o_{z_2}^2, \dots, o_{z_K}^K$  and one is left with a probability distribution

$\mathbf{P}(o_1^1, o_{z_2}^2 \dots o_{z_K}^K)$  on only  $K$  variables, one for each party. The other verifications are similar. Another way to think of it is to notice that each conditional probability factor on  $K$  variables (one variable conditioned on  $K - 1$  other variables) effectively sets a joint probability distribution on those same  $K$  variables. In the distribution (A2) there are  $\sum_{i=1}^K m_i - K + 1$  such factors and so this is exactly how many local points  $P_L(\mathbf{o}|\mathbf{z})$  that can be matched for

a given hidden variable value (see equation (20) in the main text). The argument for tightness still works if we consider only two parties among  $K$ . For any two parties we can choose a pair of inputs for each to return to a CHSH-type scenario, then the argument follows in the same way as in the proof of Lemma 1.

□